

Introducción a la Bioinformática:

Algoritmos Bioinformáticos

Luis Garreta

luis.garreta@javerianacali.edu.co

Doctorado en Ingeniería
Pontificia Universidad Javeriana – Cali

19 de marzo de 2018

Datos biológicos

- ▶ Por miles de años los humanos se han maravillado de la naturaleza de la vida: Herencia, Evolución, Enfermedades, ...
- ▶ Hace 60 años James Watson and Francis Crick sugirieron que el ADN es el material responsable para la herencia
- ▶ Esto llevó a una revitalización de los estudios de la biología con consecuencias abrumadoras
- ▶ En los últimos años, los adelantos en técnicas experimentales de biología han permitido a los científicos generar cantidades de datos sin precedentes.

Next: Introducción

El gran inicio fué el secuenciamiento del Genoma Humano

- ▶ Se tiene ya la información que existe dentro de cada célula humana: Genoma Humano.
- ▶ El entendimiento de estas instrucciones genéticas prometen permitir a los científicos un mejor entendimiento de los diferentes aspectos de la vida:
 - ▶ Enfermedades y sus tratamientos
 - ▶ Crecimiento y la vejez
 - ▶ Nuestra evolución
 - ▶ Como nos relacionamos con otras especies
 - ▶ y muchos otros aspectos

Next: Introducción

Datos biológicos y métodos computacionales

- ▶ El principal inconveniente de los investigadores y el conocimiento que buscan es la gran cantidad de datos
- ▶ Los biólogos han sido entrenados principalmente para obtener nueva información
- ▶ Y con la llegada de datos genómicos, se crearon grandes repositorios o BDs pero faltaban las herramientas para analizarlos.
- ▶ Afortunadamente, las ciencias de la computación han creado desde hace tiempo métodos, técnicas, y enfoques para analizar datos que se están aplicando para entender y mejorar la vida del hombre.

Una guía desde la ingeniería a la biología molecular

- ▶ El ácido dextribonucleico o ADN es el material genético
- ▶ En otras palabras, la información guardada en el ADN permite la organización de moléculas inanimadas en células vivientes y organismos funcionales.
 - ▶ Regulan su composición química interna, su crecimiento y reproducción
- ▶ El ADN también nos permite heredar nuestros rasgos: color de cabello, ojos, enfermedades, etc.
- ▶ Todas estas unidades que gobiernan estas características, ya sea composición química o los rasgos, se conocen como **genes**

Los Genes son secuencias de Nucleótidos

- ▶ Los genes contienen su información como secuencias específicas de nucleótidos que se encuentran en moléculas de ADN
- ▶ Solo cuatro nucleótidos diferentes (o bases) son usadas en las moléculas de ADN: Adenina (A), Guanina (G), Citocina (C), y Timina (T)
- ▶ Toda la información dentro de un gen viene dada simplemente o en que orden aparezcan estos cuatro nucleótidos.
- ▶ Un gen puede contener muchos miles de nucleótidos
- ▶ Y gran cantidad de genes codifican para proteínas y se estima que diferentes miles de estas son las que forma a un ser humano

Las proteínas son secuencias de aminoácidos

- ▶ La cantidad y el tipo de proteínas expresadas en los individuos hacen que estos sean diferentes:
 - ▶ Diferente color de ojos, de piel, o más saludables, etc.
- ▶ Una proteína es una cadena (o polímero) de bloques interconectados llamados **aminoácidos**
- ▶ Combinaciones de 20 aminoácidos (residuos) son usadas para construir toda la proteína que constituyen a un ser.

Las forma de las proteínas determina su función

- ▶ Cada uno de los 20 AA tiene propiedades que lo hacen diferente de los demás
- ▶ Algunos AA contienen átomos que tienen carga negativa (oxígeno); otros tienen átomos que tienen carga negativa (nitrógeno)
- ▶ Entonces, como imanes los átomos de los AA se atraen o repelen
- ▶ Estas atracciones/repulsiones además del medio acuoso en el que están las proteínas hacen que los AA se plieguen
- ▶ Este plegamiento le da a la proteína una forma 3D única
- ▶ Y esta forma es la que le permite realizar una función determinada dentro del organismo.

Temas del Modulo de Algoritmos en Bioinformática

- ▶ Dynamic programming: Needleman-Wunsch, Smith-Waterman, and alignment heuristics
- ▶ Phylogenetic analysis: Distance and character based phylogeny reconstruction algorithms
- ▶ Data redundancy and homology reduction: Hobohm and other clustering algorithms.
- ▶ Weight matrices: Sequence weighting, pseudo count correction for low counts, Gibbs sampling, and Psi-Blast Modelos
- ▶ Motif Discovery: Weight matrices, Sequence weighting, pseudo count correction for low counts, Gibbs sampling
- ▶ Hidden Markov Models: Model construction, Viterbi decoding, posterior decoding, and Baum Welsh HMM learning

Evaluación

Item	Valor
Tareas	30 %
Ensayos	30 %
Proyecto	40 %

Referencias

Computational books

- ▶ Fundamental Concepts of Bioinformatics (2003) by Dan E. Krane, Michael L. Raymer
- ▶ Introduction to algorithms in bioinformatics (2016, course notes) by István Miklós

Biological books

- ▶ Bioinformatics and functional genomics (2015, Third edition) by Jonathan Pevsner

Programming books

- ▶ Bioinformatics with Python Cookbook (2015) by Tiago Antao
- ▶ Bioinformatics Programming Using Python (2010) by Mitchell L Model

Sitio del Curso

Todo el material (diapositivas, lecturas, datos, código) se guardará en el repositorio público:

`https://github.com/puj2018/algoritmos-bioinformatica/`

Puede descargarlo o clonarlo:

- ▶ Para descargarlo:

`https://github.com/puj2018/algoritmos-bioinformatica/archive/master.zip`

- ▶ Para clonarlo:

`git clone https://github.com/puj2018/algoritmos-bioinformatica.git`